Sentiment Analysis

A Probabilistic Approach

S. A. Gieske   S. Laan   D. S. Ten Velthuis
C. R. Verschoor   A. J. Wiggers

Faculty of Science (FNWI)
University of Amsterdam

February 2, 2012

Outline

Outline

The goal of the project

Project Description

Performing sentiment analysis on messages about the EO

- Classification Sentiment vs. Non Sentiment
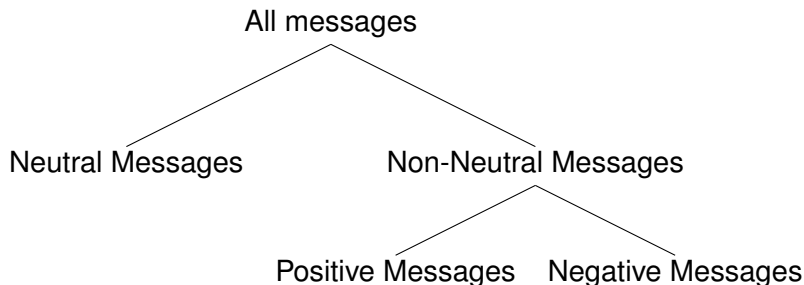- Classification Positive vs. Negative

Outline

Approach

- Preprocessing of the data
- Perform machine learning algorithms on data
- Use best algorithms to classify real time on server

Hierarchical Classification

All messages

Neutral Messages   Non-Neutral Messages

Positive Messages   Negative Messages

Outline

Outline

3 Data Preprocessing
- Dataset Analysis
- Data Cleaning
- Data Reduction

Dataset Analysis

Dataset messages EO

10.000 messages, 19 features per message

Ony 3 features used:

- Source
- Sentiment
- Message contents

Data Cleaning

Data Cleaning

- Shorten words, ex. hahaha to haha
- Stemmer

Data Reduction

- Only use Twitter messages (83% of all messages)
- Remove articles, reference words and prepositions
- Substitute smileys with words
- Remove some punctuation marks (ex. not ! ?)

Outline

Outline

Classification

Weighted Sum Probability

Weighted Sum Probability

- Create tokens
- Assign sentiment probabilities to tokens

$$P(word) = \frac{\sum word \in C_1}{\sum word \in C_1 \cup C_2} \tag{1}$$

- Assign sentiment probabilities to sentences

$$P(s) = \frac{1}{n} \sum_{w \in s} P(w) \tag{2}$$

Perceptron

Perceptron

Algorithm

Train lineair treshold

Input: sentence probabilities, sentence values
Output: treshold

Perceptron

Results & Conclusion

Results : High precision OR recall, never both

Conclusion : Lineaire threshold not good enough

Support Vector Machine

Support Vector Machine

Results & Conclusion

Results :

Conclusion :

Naive Bayes

Naive Bayes

Results & Conclusion

Results :
Conclusion :

Multiclassification with Perceptron

Multiclassification with Perceptron

Results & Conclusion

Results :

Conclusion :

The goal	Approach	Data Preprocessing	Classification	Webserver Framework	Conclusion

Entropy

Entropy

Entropy

Entropy

Results & Conclusion

Results :
Conclusion :

Neural Network

Neural Network

Results & Conclusion

Results :

Conclusion :

Outline

Webserver Framework

Request (HTML) $\rightarrow$ Server (PHP/PYTHON) $\rightarrow$ Result (XML)

Request   http://url.com/?**dataset**=1&**message**=De EO is cool!

Result   XML File (Containing: Status, Message, Sentiment, Accuracy, Precision, Recall)

Demo

Action...

Outline

Conclusion

Conclusion